数理解析研究所講究録
第311巻 1977年 109-120

予測平均2乗誤差を最小にする実験計画について

東京大 工学部　　三輪 哲久

線形回帰モデルにおける最小2乗法の理論は，モデルを仮定して，その未知パラメータに関して推論が行なわれる。しかし仮定したモデルが誤っている場合には，最小2乗推定量の不偏性とか最小分散性とかの性質が保たれない。従って実験計画の良さを論じる場合に，推定量の分散（あるいは分散共分散行列）のみに注目するのでは不十分である。

ここでは，予測平均2乗誤差を，予測を行ないたい領域にわたって重みをつけて積分したものを実験計画の良さの基準として用いる。

§1 序

線形回帰モデル

$$
\begin{aligned}
y_{ij} &= \eta(x_i) + e_{ij} = \boldsymbol{f}(x_i)'\boldsymbol{\alpha} + e_{ij} \\
&= f_1(x_i)\alpha_1 + \cdots + f_m(x_i)\alpha_m + e_{ij} \qquad (1.1)
\end{aligned}
$$

$$i=1,2,\cdots,n\ ;\ j=1,2,\cdots,r_i\ ;\ \sum_{i=1}^{n} r_i = N$$

を考える。実験点 x_i のとりうる値の範囲を実験領域と呼び，$\mathcal{X}$ で表わす。e_{ij} は互いに独立に，平均 0，分散 σ^2 の同じ分布に従う確率変数であるとする。$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_m)$ は未知のパラメータである。一般には $f(x)$ の形も未知である事が多い。従って，実際には，モデルを仮定してパラメータの推定や応答の予測を行なうのである。

予測を行なうために仮定したモデルを

$$\eta_1(x) = f_1(x)'\theta_1 \tag{1.2}$$

とする。一方 (1.1) 式を真のモデルと呼ぶ事にする。$f(x)$ のうち，$f_1(x)$ の要素でないものから成るベクトルを $f_2(x)$ とし，それに対応するパラメータを α_2 とする。

例 1
$$\begin{cases} \eta(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 , \\ \eta_1(x) = \theta_0 + \theta_1 x \end{cases}$$
$$f_1(x) = (1, x)' ,\ \theta_1 = (\theta_0, \theta_1)' ,$$
$$f_2(x) = (x^2) ,\ \alpha_2 = (\alpha_2) .$$

例 2
$$\begin{cases} \eta(x) = \alpha_0 + \alpha_1 e^{x} + \alpha_2 e^{-x} , \\ \eta_1(x) = \theta_0 + \theta_1 x + \theta_2 x^2 . \end{cases}$$
$$f_1(x) = (1, x, x^2)' ,\ \theta_1 = (\theta_0, \theta_1, \theta_2)' ,$$
$$f_2(x) = (e^{x}, e^{-x})' ,\ \alpha_2 = (\alpha_1, \alpha_2)'$$

さて，モデル (1.2) を仮定した場合の最小 2 乗推定量は

$$\hat{\theta}_1 = (X_1' X_1)^{-1} X_1' y \tag{1.3}$$

で与えられる。ただし

$$X_1' = [\underbrace{f_1(x_1) \cdots f_1(x_1)}_{r_1} \cdots \underbrace{f_1(x_n) \cdots f_1(x_n)}_{r_n}], \tag{1.4}$$

$$y' = (y_{11} \cdots y_{1r_1} \cdots y_{n1} \cdots y_{nr_n})' \tag{1.5}$$

である。また，行列

$$X_2' = [f_2(x_1) \cdots f_2(x_1) \cdots f_2(x_n) \cdots f_2(x_n)] \tag{1.6}$$

$$M_{11} = X_1' X_1 = \sum_{i=1}^{n} r_i f_1(x_i) f_1(x_i)' \tag{1.7}$$

$$M_{12} = X_1' X_2 = \sum_{i=1}^{n} r_i f_1(x_i) f_2(x_i)' \tag{1.8}$$

を定義しておく．(1.3)を使って点xにおける応答の予測は

$$\hat{y}_1(x) = f_1(x)' \hat{\theta}_1 = f_1(x)' (X_1' X_1)^{-1} X_1' y \tag{1.9}$$

であたえられる。その期待値と分散に関して

$$\begin{aligned} E[\hat{y}_1(x)] - \eta(x) &= [f_1(x)' (X_1' X_1)^{-1} X_1' X_2 - f_2(x)'] \alpha_2 \\ &= [f_1(x)' M_{11}^{-1} M_{12} - f_2(x)'] \alpha_2 \end{aligned} \tag{1.10}$$

$$V[\hat{y}_1(x)] = \sigma^2 f_1(x)' (X_1 X_1')^{-1} f_1(x) = \sigma^2 f_1(x)' M_{11}^{-1} f_1(x) \tag{1.11}$$

が成りたつ。これより予測平均2乗誤差は

$$\begin{aligned} E[\{\hat{y}_1(x) - \eta(x)\}^2] &= \sigma^2 f_1(x)' M_{11}^{-1} f_1(x) \\ &\quad + \{[f_1(x)' M_{11}^{-1} M_{12} - f_2(x)'] \alpha_2\}^2 \end{aligned} \tag{1.12}$$

となる。この値は予測を行なう点xに依存する。そこで，予測平均2乗誤差の重み付き積分

$$\int E[\{\hat{y}_1(x) - \eta(x)\}^2] w(x) dx$$

$$= \sigma^2 \operatorname{tr} M_{11}^{-1} \Lambda_{11}$$

$$+ \alpha_2' [\Lambda_{22} - 2M_{21} M_{11}^{-1} \Lambda_{12} + M_{21} M_{11}^{-1} \Lambda_{11} M_{11}^{-1} M_{12}] \alpha_2 \qquad (1.13)$$

を考える。$w(x)$は予測を行なう場所に関して，真の値からのずれに対する重要性といったものを表わす。また

$$\Lambda_{11} = \int f_1(x) f_1(x)' w(x) dx, \quad \Lambda_{12} = \int f_1(x) f_2(x)' w(x) dx,$$

$$\Lambda_{22} = \int f_2(x) f_2(x)' w(x) dx \qquad (1.14)$$

である。(1.13) を最小にするような計画

$$\epsilon(N) = \begin{Bmatrix} x_1, & x_2, & \cdots, & x_n \\ r_1, & r_2, & \cdots, & r_n \end{Bmatrix}, \quad \sum_{i=1}^{n} r_i = N \qquad (1.15)$$

を求めるのが目的であるが，ここでは連続計画

$$\epsilon = \begin{Bmatrix} x_1, & x_2, & \cdots, & x_n \\ p_1, & p_2, & \cdots, & p_n \end{Bmatrix}, \quad 0 < p_i \leqq 1, \quad \sum_{i=1}^{n} p_i = 1 \qquad (1.16)$$

を考える。

$$M_{11} = N \sum_{i=1}^{n} \frac{r_i}{N} f_1(x_i) f_1(x_i)' = N \sum_{i=1}^{n} p_i f_1(x_i) f_1(x_i)' = N M_{11}(\epsilon) \quad (1.17)$$

$$M_{12} = N \sum_{i=1}^{n} \frac{r_i}{N} f_1(x_i) f_2(x_i)' = N \sum_{i=1}^{n} p_i f_1(x_i) f_2(x_i)' = N M_{12}(\epsilon) \quad (1.18)$$

を用いて(1.13)式を書きなおすと

$$\int E[\{\hat{y}_1(x) - \eta(x)\}^2] w(x) dx$$

$$= \frac{\sigma^2}{N} \operatorname{tr} M_{11}^{-1}(\epsilon) \Lambda_{11}$$

$$+ \alpha_2' [\Lambda_{22} - M_{21}(\epsilon) M_{11}^{-1}(\epsilon) \Lambda_{12} + M_{21}(\epsilon) M_{11}^{-1}(\epsilon) \Lambda_{11} M_{11}^{-1}(\epsilon) M_{12}(\epsilon)] \alpha_2$$

$$= V_1(\epsilon) + B_1(\epsilon) = H_1(\epsilon) \quad . \tag{1.19}$$

(1.13) を最小にする計画は, Box and Draper [1, 2] が特別な場合について考えている。そこでは, 真のモデルも推定のために仮定したモデルも多項式であり, 実験領域は原点のまわりの超球, あるいは超立方体に限られている。さらに, 最適な計画は限定された集合の中から求められている。本稿で述べる議論は, そのような制約をすべて取り去ったより広範な場合に適用できる。

§2　$H_1(\epsilon)$を最小にする実験計画

$H_1(\epsilon)$の値は真のモデルが何であるかと言う事と, そのパラメータの値に依存する。従って$H_1(\epsilon)$を最小にするという問題は人工的なものである。しかし最適な計画の性質を調べることは実際に実験を行なう際に, 少なからず役立つ情報を与える。まず, $H_1(\epsilon)$を最小にする計画を特徴づける定理を述べる。

<u>定理</u>　次のような4つの条件を考える。

(1)　計画ϵ^*は$H_1(\epsilon)$を最小にする.

(2)　計画ϵ^*は $\max_{x \in \mathscr{X}} \varphi_1(x;\epsilon)$ を最小にする。ただし

$$\begin{aligned}\varphi_1(x;\epsilon) = B_1(\epsilon) + \frac{\sigma^2}{N} \operatorname{tr} M_{11}^{-1}(\epsilon) f_1(x) f_1(x)' M_{11}^{-1}(\epsilon) \Lambda_{11} \\ + 2\alpha_2' [f_2(x) f_1(x)' - M_{21}(\epsilon) M_{11}^{-1}(\epsilon) f_1(x) f_1(x)'] \\ \times M_{11}^{-1}(\epsilon) [\Lambda_{12} - \Lambda_{11} M_{11}^{-1}(\epsilon) M_{12}(\epsilon)] \alpha_2 \quad .\end{aligned} \tag{2.1}$$

(3) $\max\limits_{x \in \mathcal{X}} \varphi_1(x ; \overset{*}{\epsilon}) = H_1(\overset{*}{\epsilon})$ 。 $\quad$ (2.2)

(1′) 任意の計画 ϵ に対して $\tilde{\epsilon} = (1-\alpha)\overset{*}{\epsilon} + \alpha \epsilon$ とすると

$$\frac{\partial}{\partial \alpha} H_1(\tilde{\epsilon})\Big|_{\alpha=0} \geqq 0 \quad . \tag{2.3}$$

このとき，次のような必要，十分の関係が成りたつ。

$$\begin{tikzcd} (1) \arrow[d, shift right] \arrow[drr] & & (1') \\ (2) \arrow[u, shift right] & & (3) \arrow[u] \end{tikzcd} \tag{2.4}$$

定理の証明を述べる前に，補題を２つ挙げる。

<u>補題１</u>　任意の計画 ϵ_1，ϵ_2 に対して $\epsilon = (1-\alpha)\epsilon_1 + \alpha \epsilon_2$ とする。

$$\begin{aligned} \frac{\partial}{\partial \alpha} H_1(\epsilon)\Big|_{\alpha=0} = {} & \frac{\sigma^2}{N} \operatorname{tr} M_{11}^{-1}(\epsilon_1)\Lambda_{11} - \frac{\sigma^2}{N} \operatorname{tr} M_{11}^{-1}(\epsilon_1) M_{11}(\epsilon_2) M_{11}^{-1}(\epsilon_1)\Lambda_{11} \\ & - 2\alpha_2' \left[M_{21}(\epsilon_2) - M_{21}(\epsilon_1) M_{11}^{-1}(\epsilon_1) M_{11}(\epsilon_2) \right] M_{11}^{-1}(\epsilon_1) \\ & \times \left[\Lambda_{12} - \Lambda_{11} M_{11}^{-1}(\epsilon_1) M_{12}(\epsilon_1) \right] \alpha_2 \quad . \end{aligned} \tag{2.5}$$

<u>証明</u>　一般に $\frac{\partial}{\partial \alpha} A^{-1} = -A^{-1} \frac{\partial A}{\partial \alpha} A^{-1}$ であることと，$M_{11}(\epsilon) = (1-\alpha) M_{11}(\epsilon_1) + \alpha M_{11}(\epsilon_2)$，$M_{12}(\epsilon) = (1-\alpha) M_{12}(\epsilon_1) + \alpha M_{12}(\epsilon_2)$ である事を用いれば証明できる。

<u>補題２</u>　任意の計画 $\epsilon = \left\{ \begin{matrix} x_1, & x_2, & \cdots, & x_n \\ p_1, & p_2, & \cdots, & p_n \end{matrix} \right\}$ に対して

(1) $\sum\limits_{i=1}^{n} p_i \varphi_1(x_i ; \epsilon) = H_1(\epsilon)$， $\quad$ (2.6)

(2) $\max\limits_{x \in \mathcal{X}} \varphi_1(x ; \epsilon) \geqq H_1(\epsilon)$ 　. $\quad$ (2.7)

<u>証明</u>　$\sum\limits_{i=1}^{n} p_i f_1(x_i) f_1(x_i)' = M_{11}(\epsilon)$，$\sum\limits_{i=1}^{n} p_i f_1(x_i) f_2(x_i)' = M_{12}(\epsilon)$ であることを使えば

$$\sum_{i=1}^{n} p_i \varphi_1(x_i ; \epsilon) = B_1(\epsilon) + \frac{\sigma^2}{N} \operatorname{tr} M_{11}^{-1}(\epsilon)\Lambda_{11} = H_1(\epsilon) \tag{2.10}$$

が成りたつ。さらに

$$\max_{x\in\mathscr{X}} \varphi_1(x;\epsilon) = \max_{x\in\mathscr{X}} \varphi_1(x;\epsilon)\sum_{i=1}^{n} p_i$$
$$\geq \sum_{i=1}^{n} p_i\,\varphi_1(x_i;\epsilon) = H_1(\epsilon) \qquad (2.11)$$

が成りたつ。　　　　　　　　　　　　　　　証明終。

<u>定理の証明</u>　まず，(1)と(2)の同値性を示す。$H_1(\epsilon)$を最小にする計画を$\overset{*}{\epsilon}$とする。任意の計画ϵに対して$\tilde{\epsilon}=(1-\alpha)\overset{*}{\epsilon}+\alpha\epsilon$を作ると，$\overset{*}{\epsilon}$の定義より

$$\frac{\partial}{\partial\alpha} H_1(\tilde{\epsilon})\Big|_{\alpha=0} \geq 0 \qquad (2.12)$$

である。ϵとして1点xのみから成る計画$\epsilon(x)$を考えると

$$M_{11}[\epsilon(x)] = f_1(x)f_1(x)',\quad M_{12}[\epsilon(x)] = f_1(x)f_2(x)'. \qquad (2.13)$$

補題1において，$\epsilon_1=\overset{*}{\epsilon}$，$\epsilon_2=\epsilon(x)$　を代入すれば，

$$\frac{\partial}{\partial\alpha} H_1(\tilde{\epsilon})\Big|_{\alpha=0} = H_1(\overset{*}{\epsilon}) - \varphi_1(x;\overset{*}{\epsilon}) \geq 0 \qquad (2.14)$$

従って

$$\max_{x\in\mathscr{X}} \varphi_1(x;\overset{*}{\epsilon}) \leq H_1(\overset{*}{\epsilon}). \qquad (2.15)$$

一方，$\max_{x\in\mathscr{X}} \varphi_1(x;\epsilon)$を最小にする計画を$\overset{\circ}{\epsilon}$とする。定義より

$$H_1(\overset{*}{\epsilon}) \leq H_1(\overset{\circ}{\epsilon}), \qquad (2.16)$$

$$\max_{x\in\mathscr{X}} \varphi_1(x;\overset{\circ}{\epsilon}) \leq \max_{x\in\mathscr{X}} \varphi_1(x;\overset{*}{\epsilon}). \qquad (2.17)$$

また，補題2より

$$H_1(\overset{\circ}{\epsilon}) \leq \max_{x\in\mathscr{X}} \varphi_1(x;\overset{\circ}{\epsilon}). \qquad (2.18)$$

従って

$$H_1(\overset{*}{\epsilon}) \leq H_1(\overset{\circ}{\epsilon}) \leq \max_{x\in\mathscr{X}} \varphi_1(x;\overset{\circ}{\epsilon}) \leq \max_{x\in\mathscr{X}} \varphi_1(x;\overset{*}{\epsilon}) \leq H_1(\overset{*}{\epsilon}) \qquad (2.19)$$

であるから(2.19)においては，すべて等号が成りたつ。すなわち条件(1)と(2)は同値である。また(2.19)より，(1)あるいは(2)が成りたてば，(3)が成りたつことも明らかである。

次に(1)が(3)の必要条件である事を示す。(3)をみたす計画を $\bar{\epsilon}$ とする。すなわち

$$\max_{x \in \mathscr{X}} \varphi_1(x;\bar{\epsilon}) = H_1(\bar{\epsilon}) \quad。 \tag{2.20}$$

任意の計画 $\epsilon = \sum_{i=1}^{n} p_i \, \epsilon(x_i)$ に対して，$\tilde{\epsilon} = (1-\alpha)\bar{\epsilon} + \alpha\epsilon$ として補題1を使えば

$$\begin{aligned}\frac{\partial}{\partial\alpha} H_1(\tilde{\epsilon})\Big|_{\alpha=0} &= H_1(\bar{\epsilon}) - \sum_{i=1}^{n} p_i \, \varphi_1(x_i;\bar{\epsilon}) \\ &\geq H_1(\bar{\epsilon}) - \max_{x \in \mathscr{X}} \varphi_1(x;\bar{\epsilon}) \sum_{i=1}^{n} p_i = 0 \ . \end{aligned} \tag{2.21}$$

すなわち(1)が成りたつ。 証明終。

§3．例

真のモデル，及び予測に使うモデルが

$$\eta(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 \tag{3.1}$$

$$\eta_1(x) = \theta_0 + \theta_1 x \tag{3.2}$$

であるとする。実験領域は $\mathscr{X} = [-1, 1]$ であり，重みは $w(x) = 1 \quad (-1 \leq x \leq 1)$ とする。

(i) $\overline{V}_1(\epsilon)$ を最小にする計画 $\overset{*}{\epsilon}_V$.

$$\overset{*}{\epsilon}_V = \begin{Bmatrix} -1 & , & 1 \\ \frac{1}{2} & & \frac{1}{2} \end{Bmatrix} \tag{3.3}$$

$$H_1(\overset{*}{\epsilon}_V) = \frac{8}{3}\frac{\sigma^2}{N} + \frac{15}{16}\alpha_2^2 \tag{3.4}$$

(ii)　$B_1(\epsilon)$を最小にする計画$\overset{*}{\epsilon}_B$.

$$\overset{*}{\epsilon}_B = \begin{Bmatrix} -1 & , & 0 & , & 1 \\ 1/6 & , & 2/3 & , & 1/6 \end{Bmatrix} \tag{3.5}$$

$$H_1(\overset{*}{\epsilon}_B) = 4\frac{\sigma^2}{N} + \frac{8}{45}\alpha_2^2 \tag{3.6}$$

(iii)　$H_1(\epsilon)$を最小にする計画$\overset{*}{\epsilon}$。

$$\overset{*}{\epsilon} = \begin{Bmatrix} -1 & , & 0 & , & 1 \\ \overset{*}{p} & , & 1-2\overset{*}{p} & , & \overset{*}{p} \end{Bmatrix} \tag{3.7}$$

最適計画の$\overset{*}{p}$の値はα_2の値に依存する。表１には$\beta_2 = \dfrac{\alpha_2}{\sigma/\sqrt{N}}$の種々の値に対する$\overset{*}{p}$の値が示されている。また，$V_1(\overset{*}{\epsilon})$，$B_1(\overset{*}{\epsilon})$，$H_1(\overset{*}{\epsilon})$，$H_1(\overset{*}{\epsilon}_V)$，$H_1(\overset{*}{\epsilon}_B)$の値を$\sigma^2/N$で割った値も示されている。図１は，各値をプロットしたものである。

(iv)　モデル$\eta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$　を予測に使うと，

$$H(\epsilon) = V(\epsilon) \tag{3.8}$$

であり，最適計画は

$$\overset{*}{\epsilon} = \begin{Bmatrix} -1 & , & 0 & , & 1 \\ 1/4 & , & 1/2 & , & 1/4 \end{Bmatrix} \tag{3.9}$$

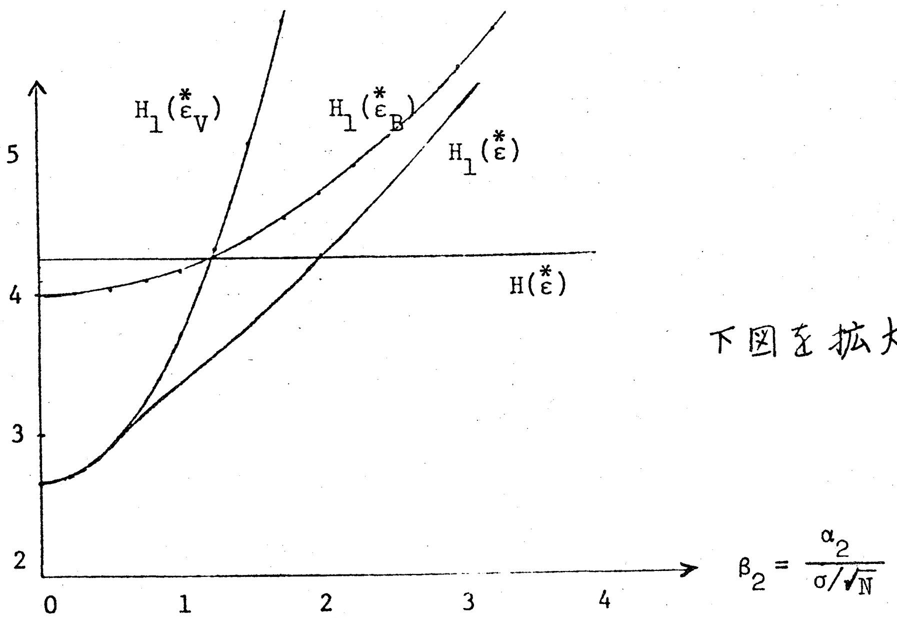
である。$H(\overset{*}{\epsilon})/\frac{\sigma^2}{N} = 4.267$で，図１の水平線がこれである。

§4　逐次的構成法

前節で述べたような簡単な例を除いて，一般には$H_1(\epsilon)$を最

表 1.

β_2	$\overset{*}{p}$	$V_1(\overset{*}{\varepsilon})$	$B_1(\overset{*}{\varepsilon})$	$H_1(\overset{*}{\varepsilon})$	$H_1(\overset{*}{\varepsilon}_V)$	$H_1(\overset{*}{\varepsilon}_B)$
0.0	0.5	2.667	0.0	2.667	2.667	4.0
0.1	0.5	2.667	0.011	2.677	2.677	4.002
0.2	0.5	2.667	0.043	2.709	2.709	4.007
0.3	0.5	2.667	0.096	2.763	2.763	4.016
0.4	0.5	2.607	0.171	2.837	2.837	4.028
0.5	0.5	2.607	0.267	2.933	2.933	4.044
0.6	0.451	2.739	0.297	3.036	3.051	4.064
0.7	0.414	2.805	0.328	3.132	3.189	4.087
0.8	0.386	2.864	0.359	3.224	3.349	4.114
0.9	0.362	2.920	0.393	3.312	3.531	4.144
1.0	0.343	2.971	0.428	3.398	3.733	4.178
2.0	0.250	3.333	0.933	4.267	6.933	4.711
3.0	0.216	3.542	1.777	5.318	12.267	5.600
4.0	0.199	3.672	2.982	6.653	19.733	6.844
5.0	0.190	3.756	4.551	8.308	29.333	8.444
6.0	0.184	3.814	6.485	10.298	41.067	10.400
7.0	0.180	3.854	8.779	12.633	54.933	12.711
8.0	0.177	3.883	11.433	15.316	70.933	15.378
9.0	0.175	3.904	14.446	18.310	89.067	18.400
10.0	0.174	3.920	17.816	21.736	109.333	21.778



下図を拡大したもの

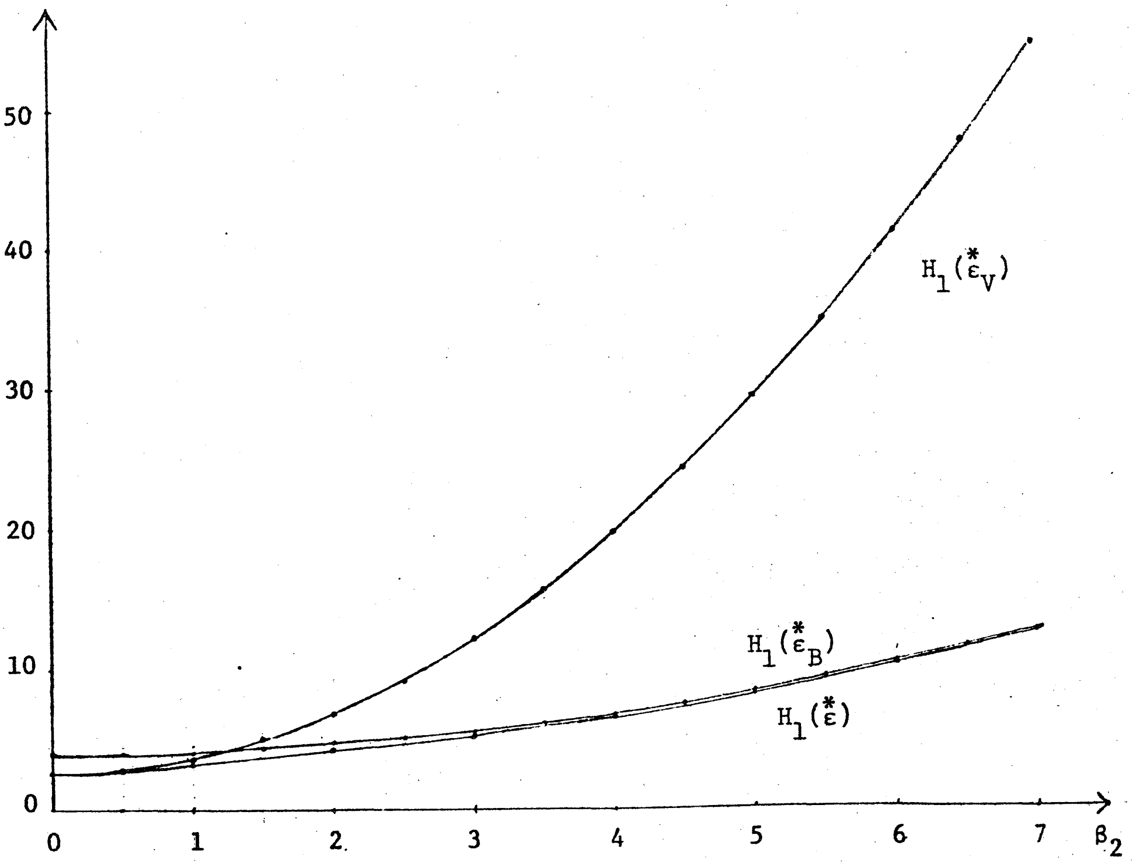


図　1

小にする計画を解析的に求めることは困難である。ここでは,その逐次的な構成法について述べる。

任意の計画 ϵ_0 と1点より成る計画 $\epsilon(x)$ に対して

$$\epsilon_1 = (1-\alpha)\,\epsilon_0 + \alpha\,\epsilon(x) \tag{4.1}$$

とすると

$$H_1(\epsilon_1) \simeq H_1(\epsilon_0) + \alpha\left[\frac{\partial}{\partial\alpha} H_1(\epsilon_1)\right]_{\alpha=0}$$

$$= H_1(\epsilon_0) + \alpha\,[\,H_1(\epsilon_0) - \varphi_1(x;\epsilon_0)\,]\,. \tag{4.2}$$

もし ϵ_0 が定理の(i)を満たさないならば,定理と補題2より $\varphi_1(x_0;\epsilon_0) > H_1(\epsilon_0)$ なる点 x_0 が存在する。従って十分小さい α を選べば $H_1(\epsilon_1) < H_1(\epsilon_0)$ とできる。

逐次的構成法

1. $M_{11}(\epsilon_s)$ が正則であるような計画 ϵ_s を与える。
2. $\varphi_1(x_s;\epsilon_s) = \max_{x\in\mathcal{X}} \varphi_1(x;\epsilon_s) > H_1(\epsilon_s)$ なる点 x_s を求める。
3. $\epsilon_{s+1} = (1-\alpha_s)\,\epsilon_s + \alpha_s\,\epsilon(x_s)$.

この手順は定理の(i)を満たす計画に収束する。

参考文献

[1] Box,G.E.P. & Draper,N.R.(1959). A basis for the selection of a responce surface design. *J. Amer. Statist. Ass.*, Vol.54, 622-654.

[2] Box,G.E.P. & Draper,N.R.(1963). The choice of a second order rotatable design. *Biometrika*. Vol.50, 335-352.